Penggerombolan Model Parameter Regresi dengan *Error-Based Clustering*

[1]I Made Sumertajaya
[2]Gusti Adhi Wibawa
[3]I Gede Nyoman Mindra Jaya

[1]Staf Pengajar Departemen Statistika IPB
[2,3]Mahsiswa Pascasarjana Statistika IPB

ABSTRAK

Ketersediaan data tidak dalam format standar yaitu tidak dalam bentuk vektor dalam dimensi ruang p, sering kali menjadi kendala dalam penggunaan analisis gerombol tradisonal. Untuk dapat menggunakan teknik analisis gerombol tradisional, data terlebih dahulu harus dirubah ke dalam struktur yang diinginkan untuk mempermudah analisis. Namun tidak jarang dalam proses mengubah struktur data awal menjadi struktur data baru banyak informasi yang hilang. Dalam setiap teknik ini disajikan statistik varians-kovarians atau matriks kekeliruan (measurement Error) yang terkait dengan hasil perubahan struktur data tersebut yang merupakan ukuran informasi yang hilang selama proses tranformasi. Metode Error based clustering memungkinkan melakukan penggerombolan objek dengan emmperhatikan kekeliruan pengukuran. Salah satu aplikasi dari metode ini adalah penggerombolan parameter regresi dalam kasus klasifikasi sekuritas dalam perdagangan saham.

Key Word : *Error-Based Clustering, kError*

PENDAHULUAN

Teknik analisis gerombol adalah suatu proses pengorganisasian data yang sangat besar kedalam kelompok-kelompok yang lebih kecil dengan data yang memiliki kemiripan ditempatkan pada kelompok yang sama sedangkan yang kurang mirip ditempakan dalam kelompok yang berbeda. Teknik ini diharapkan mampu mengungkapkan informasi yang tersimpan dalam data sehingga bisa membantu dalam pengambilan keputusan yang tepat terkait dengan kajian yang sedang dilakukan.

Pemanfaatan data yang tidak standar yaitu tidak dalam bentuk vektor dalam dimensi ruang *p*, sering kali menjadi kendala dalam penggunaan analisis gerombol tradisonal. Untuk dapat menggunakan teknik analisis gerombol tradisional, data terlebih dahulu harus dirubah ke dalam struktur yang diinginkan untuk mempermudah analisis. Namun tidak jarang dalam proses mengubah struktur data awal menjadi struktur data baru banyak informasi yang hilang. Beberapa contoh teknik yang digunakan untuk

mengubah struktur data yaitu menghitung rataan, mereduksi data dengan teknik analisis komponen utama ataupun menggunakan model-model statistik yang lain. Dalam setiap teknik ini disajikan statistik varians-kovarian atau matriks kekeliruan (*measurement Error*) yang terkait dengan hasil perubahan struktur data tersebut yang merupakan ukuran informasi yang hilang selama proses tranformasi. Hasil transformasi ini kemudian dijadikan unit pengamatan baru yang digerombolkan tanpa memperhatikan matriks kekeliruan yang menyertai setiap unit pengamatan baru tersebut.

Analisis gerombol klasik yang didasarkan pada jarak kuadrat *Euclidean* memiliki kelemahan yaitu tidak mempertimbangkan adanya informasi kekeliruan ataupun unsur ketidakpastian yang terkait dengan data. Memasukkan informasi kekeliruan dalam proses penggerombolan akan memberikan hasil pengelompokkan yang berbeda dan tentunya lebih baik dibandingkan dengan analisis gerombol tradisional seperti K-*means* dan Ward's *hierarchical clustering*.

Salah satu pendekatan baru dalam metode penggerombolan yang memasukkan informasi kekeliruan yang terkait dengan data adalah metode *Error-based clustering* (Kumar, 2007). Dalam metode *Error-based clustering* dikembangkan model statistik dan algoritma penggerombolan yang melibatkan *measurement Error.*

Terdapat dua algoritma penggerombolan dalam *Error-based clustering* yaitu (1) *hError*, yaitu algoritma pengelompokkan hirarki yang menghasilkan sebuah rangkaian gerombol tersarang, (2) *kError*, yaitu algoritma partisi, yang mempartisi data menjadi beberapa gerombol spesifik.

Salah satu aplikasi metode ini adalah penggeromobolan parameter model regresi yang dapat diterapkan pada kasus pengerombolan sekuritas saham. Model yang umumnya digunakan dalam kajian sekuritas adalah *Capital Asset Pricing Model* (CAPM).

TUJUAN PENELITIAN

Menerapkan teknik penggerombolan *Error-based clustering* khususnya k*Error* dalam penggerombolan model parameter regresi.

MODEL *ERROR-BASED CLUSTERING*

Data yang akan dianalisis terdiri dari n observasi, $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ (vector kolom) dalam ruang dimensi p (R^p) dan n matriks definit positif $\boldsymbol{\Sigma}_1,\ldots, \boldsymbol{\Sigma}_n$ dalan R^{pxp}, dimana $\boldsymbol{x}_i$ variabel pengukur dengan p karaketeristik dan $\boldsymbol{\Sigma}_i$ menyatakan matriks varians-kovarians yang terkait dengan nilai observasi $\boldsymbol{x}_i$. Misalkan bahwa setiap data independent dan dibangun dari p-variate distribusi normal dengan G kemungkinan rataan $\theta_1,...,\theta_G$, $G \le n$, dengan $x_i \sim N_p(\mu_i, \Sigma_i)$ dengan $\mu_i \in \{\theta_1,...,\theta_G\}$ untuk $i = 1, ..., n$. Tujuan dari *Error-based clustering* adalah menemukan $C_1,...,C_n$ sedemikian sehingga observasi yang memiliki nilai rataan (μ_i) yang sama masuk ke dalam gerombol yang sama sehingga $\mu_i = \theta_k$, k = 1, 2,...,G. Terdapat perbedaan dalam model penggerombolan *Error–based clustering* dengan model peluang yang lain yaitu penggerombolan secara tradisional mengasumsikan bahwa kedua nilai μ_i dan Σ_i tidak diketahui, disini kita mengasumsikan bahwa Σ_i diketahui dan μ_i yang tidak diketahui.

Misalkan $S_k = \{i | x_i \in C_k\}$ k=1,2,...,G. Perhatikan bahwa $S_1,...,S_G$ saling bebas dan bukan merupakan himpunan kosong dari $\{1,...,n\}=\bigcup_{k=1}^{G} S_k$. Sehingga $\mu_i = \theta_k$ untuk $\forall_i \in S_k$, k=1,...,G. Diberikan data pengamatan $\boldsymbol{x}_1,...,\boldsymbol{x}_n$ dan matriks kekeliruan $\boldsymbol{\Sigma}_1,..., \boldsymbol{\Sigma}_n$, prosedur fungsi kemungkinan maksimum dengan memilih $S=(S_1,...,S_G)$ dan $\theta=(\theta_1,...,\theta_G)$ sedemikian sehingga memaksimumkan fungsi likelihoodnya. Fungsi kemungkinan maksimumnya adalah:

$$L(x|S,\theta(= \prod_{k=1}^{G}\prod_{i\in S_k} \frac{1}{(2\pi)^{\frac{p}{2}}}|\Sigma_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(x_i-\theta_k)^t \Sigma_i^{-1}(x_i-\theta_k)}, \quad (1)$$

Dengan $|\boldsymbol{\Sigma}_i|$ determinan dari $\boldsymbol{\Sigma}_i$ untuk i=1,...,G

Lemma 1. Penduga fungsi kemungkinan maksimum dari $S_1,...,S_g$ adalah partisi dari n observasi ke dalam G gerombol sehingga diperoleh penyelesaian :

$$\min_{S_1,...,S_G} \sum_{k=1}^{G}\sum_{i\in S_k} (x_i - \hat{\theta}_k)^t \Sigma_i^{-1}(x_i - \hat{\theta}_k), \quad (2)$$

Dimana $\hat{\theta}_k$ adalah penduga kemungkinan maksimum dari θ_k yang diberikan oleh :

$$\hat{\theta}_k = (\sum_{i\in S_k} \Sigma_i^{-1})^{-1}(\sum_{i\in S_k} \Sigma_i^{-1} x_i), \; k = 1,...,G \quad (3)$$

Peminimuman ini memumculkan pemikiran bahwa setiap data dibobot oleh kebalikan matriks varians-kovarians kekeliruan, sehingga data dengan kekeliruan kecil akan

memiliki bobot yang lebih tinggi. Perhatikan bahwa $\hat{\theta}_k$ adalah bobot rata-rata dari data pada gerombol C_k. Jika dikaitkan dengan ukuran jarak Mahalanobis, maka $\hat{\theta}_k$ adalah rataan Mahalanobis untuk gerombol C_k. Misalkan ψ_k adalah matriks kekeliruan *p*x*p* dari yang terkait dengan $\hat{\theta}_k$, maka ψ_k dapat dituliskan sebagai berikut :

$$\psi_k = Cov(\hat{\theta}_k) = [\sum_{i \in S_k} \Sigma_i^{-1}]^{-1} \quad (4)$$

Terdapat dua hal yang menarik dari fungsi tujuan pada persamaan 2. Pertama, ketika matriks kekeliruan merupakan matriks diagonal $\Sigma_i=\sigma^2 I$, dengan bentuk *spherical*, fungsi tujuan *Error-based clustering* akan sama dengan fungsi tujuan pada algoritma *K*-Means yaitu meminimumkan jumlah kuadrat jarak *euclidean*. Kedua, fungsi tujuan dari *Error-based clustering* invarians skala seperti pada jarak Malahanobis.

ALGORITMA *kERROR*

Algoritma *kError* adalah algoritma dalam *Error-based clustering* jika jumlah gerombol G sudah diteteapkan. Konsep dasar algoritma *kError* sama dengan algoritma K-Means. Algoritma ini merupakan algoritma iterative yang mengikuti dua langkah yaitu :

Langkah 1. Untuk banyaknya gerombol yang diberikan, hitung pusat penggerombolan sebagai rata-rata jarak mahalanobis dari gerombol

Langkah 2. Perlakukan kembali setiap data untuk pusat gerombol terdekat menggunakan formulasi jarak pada persamaan 5.

Jarak data x_i dari pusat gerombol $\hat{\theta}_k$ untuk gerombol C_k adalah :

$$d_{ik} = (x_i - \hat{\theta}_k)^t \Sigma_i^{-1} (x_i - \hat{\theta}_k) \quad (5)$$

Algoritma *kError*

Algorithma 2 : *kError* (x, Σ, G)

---

1. Input : (x_i , Σ_i, G), i = 1, …, n
2. Output : Cluster C_1,…,C_G
3. Initialization :
4. Temukan inisial partisi secara acak dari data ke dalam G gerombol.
5. End initialization;

---

6. Step 1
7. Hitung pusat gerombol G dengan mengunakan persamaan 3
8. End Step 1
9. Step 2
10. Tempatkan data ke dalam gerombol terdekat menggunakan fungsi jarak pada 5
11. End step 2 :
12. If gerombol berubah then
13. Go to step 1;
14. End if
15. Kembali ke $C_1,..,C_G$

PENGGEROMBOLAN PARAMETER MODEL REGRESI

Salah satu aplikasi penggerombolan parameter model adalah penggerombolan parameter dalam model analisis regresi. Misalkan terdapat *n* objek yang diukur oleh set parameter yang diperoleh dari metode kuadrat terkecil. Misalkan terdapat m_i observasi untuk objek ke-*i* dengan *i*=1,2...,*n*. Asumsikan bahwa observasi ke-*i* mengikuti model linier sebagai berikut :

$$Y_i = X_i \beta_i + \varepsilon_i, \tag{6}$$

Dengan Y_i adalah vektor dengan m_i observasi, X_i adalah matriks dengan elemen yang telah diketahui dan berukuran m_i *x p*, β_i adalah vektor dari *p* varameter yang tidak diketahui, dan ε_i adalah vektor kekeliruan dari m_i observasi dengan rataan 0, dan varians σ^2. Pendugaan metode kuadrat terkecil memberikan dugaan β_i sebagai berikut :

$$b_i = (X_i^t X)^{-1} X_i^t Y_i, \quad i = 1,...,n, \tag{7}$$

yang meninimumkan jumlah kuadrar residual sebagai berikut :

$$SSE_i = (Y_i - X_i b_i)^t (Y_i - X_i b_i) \tag{8}$$

untuk semua pilihan b_i.

Sedangkan matriks varians kovarians yang terkait dengan b_i adalah :

$$\Sigma_i = \sigma^2 (X_i^t X_i)^{-1} \tag{9}$$

dengan σ^2 diduga sebagai berikut :

$$\hat{\sigma}^2 = \frac{(Y_i - X_i b_i)^t (Y_i - X_i b_i)}{m_i - p} \tag{10}$$

Asumsikan bahwa objek yang akan dikelompokkan digolongkan dengan baik oleh koefisien regresi dugaan sehinga kesamaan dalam koefisen regresi menunjukkan kesamaan dalam objek sesungguhnya, atau dengan kata lain, objek yang sama memiliki koefisien regresi yang sama.

Dalam kasus ini, input penggerombolan terdiri dari n set koefisien regresi dugaan b_i, dan terkait dengan matriks kekeliruan Σ_i, untuk i=1,...,n. Tujuannya adalah mempartisi b_i ke dalam G gerombol sehingga b_i yang relatif sama berada dalam gerombol yang sama, sehingga C_k memiliki dugaan parameter β_i yang sama.

Misalkan gerombol C_k berisi n_k objek dengan indeks $S_k=\{i_1, i_2,..., i_{nk}\}$. Sehingga dapat dituliskan :

$$Y_{Sk} = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_k} \end{pmatrix}, \quad X_{Sk} = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{in_k} \end{pmatrix}, \tag{11}$$

Y_{Sk} vektor berdimensi $\sum_{j=1}^{n_k} m_{ij}$ dan X_{Sk} adalah matriks berdimensi $\sum_{j=1}^{n_k} m_{ij} xp$.

Sehingga :

$$\sum_{i \in S_k} X_i^t X_i = X_{Sk}^t X_{Sk} \tag{12}$$

$$\sum_{i \in S_k} X_i^t Y_i = X_{Sk}^t Y_{Sk} \tag{13}$$

$$\sum_{i \in S_k} Y_i^t Y_i = Y_{Sk}^t Y_{Sk} \tag{14}$$

Misalkan β_{Sk} adalah parameter regresi bersama untuk objek dalam C_k, kemudian pendugaan dengan metode kuadrat terkecil diberikan sebagai berikut :

$$b_{S_k} = (X_{S_k}^t X)^{-1} X_{S_k}^t Y_{S_k} \tag{15}$$

Sedangkan matriks varians-kovarins dari dugaan parameter regresinya adalah

$$\Sigma_{S_k} = \sigma^2 (X_{S_k}^t X_{S_k})^{-1} \tag{16}$$

CONTOH APLIKASI
PENGGEROMBOLAN MODEL PARAMETER REGRESI

Untuk aplikasi *Error based clustering* pada model regresi linier multiple, akan diambil sebuah masalah penggerombolan yang sering terjadi dalam kasus bisnis saham.

Misalkan akan dikelompokkan portofolio berdasarkan pada kesamaan *performance* nya dibandingkan dengan seluruh *performance* market yang ada. Biasanya digunakan model untuk mengukur performance portofolio dibandingkan performance market yaitu dengan model Capital Asset Pricing Model (CAPM)

$$R_{it} - R_f = \alpha_i + \beta_i(R_{mt} - R_f) + \varepsilon_{it}, \text{ i=1,...,N} \quad \text{t=1,...,T} \qquad (17)$$

dengan :

R_{it} = Tingkat keuntungan sekuritas ke-i pada periode ke-t

R_f = Tingkat keuntungan bebas risiko investasi

R_{mt} = Tingkat keuntungan sekuritas pasar pada period e ke-t

β_i = Koefisien beta untuk sekuritas i yang mengukur risiko sekuritas

α_i = Ukuran kebaikan nilai portofolio sebenarnya dibandingkan dengan nilai prediksi.

Persamaan di atas juga dapat ditulis dalam bentuk

$$R_{it\,f} = \alpha_i^* + \beta_i R_{mt} + \varepsilon_{it} \qquad (18)$$

Dengan $\alpha_i^* = \alpha_i + R_f - \beta_i * R_f$

parameter regresi dapat diduga dengan menggunakan metode kuadrat terkecil dengan setiap dugaan akan disertai matriks covariansnya yang merupakan matriks kekeliruan. Untuk membuktikan bahwa teknik penggerombolan k*Error* lebih baik dibandingkan dengan metode penggerombolan k-means baik untuk data tidak distandarkan terhadap satuan pengukuran juga data distandarkan terhadap satuan pengukuran, Kumar (2007) telah melakukan simulasi dengan hasil sebagai berikut :

Tabel 1. Perbandingan Kesalahan Klasifikasi Metode Penggerombolan

Metode Penggerombolan	Rata-rata Kesalahan Klasifikasi
k*Error*	0%
k-Means	8.53%
k-Means dengan standardisasi	5.31%

Untuk menjelaskan proses perhitungan dan menunjukkan perbedaan dari k*Error* dengan k-Means, penulis akan meggerombolkan portofolio berdasarkan pada kesamaan *performance* nya dibandingkan dengan performance pasar. Sembilan perusahaan yang

termasuk dalam LQ-45 diambil sebagai sampel. Data return Saham dan return pasar ditunjukkan dalam tabel di bawah ini .

Tabel 2. Data Return Saham 9 Perusahaan dan Return Pasar

Bulan	Return Saham									Return Pasar
	Rif(A)	Rif(B)	Rif(C)	Rif(D)	Rif(E)	Rif(F)	Rif(G)	Rif(H)	Rif(I)	Rm
Juni	26.34	10.16	29.41	8.85	15.45	2.46	25.14	22.38	22.50	13.38
Juli	-5.83	-12.04	-9.09	-0.33	5.12	0.64	-6.25	16.57	1.22	-4.45
Agustus	-6.34	-17.77	8.50	-5.71	-7.49	6.04	-1.91	7.23	-8.07	-5.24
September	-15.99	-4.36	-9.22	-3.81	-8.87	0.75	-15.05	1.48	-20.18	-9.66
Oktober	-4.15	-5.97	4.06	-13.67	-2.27	10.12	2.86	7.30	-3.85	-3.80
Nopember	31.34	34.33	-2.44	2.08	27.73	6.08	9.44	-14.97	5.71	5.89
Desember	-1.52	0.00	0.00	-16.33	6.05	0.00	6.60	0.00	-16.22	-3.00
Januari	-0.39	-11.11	-15.00	18.29	-20.47	-0.64	2.86	25.20	80.65	-8.47
Februari	7.72	31.25	0.00	20.62	20.25	-0.64	2.78	7.03	3.57	12.40
Maret	-7.17	-23.81	0.00	-17.09	-16.84	0.65	-2.70	-6.57	-3.45	-11.03
April	-11.58	-0.63	-8.82	-1.03	5.06	0.00	-2.78	-2.24	-30.36	-5.99
Mei	6.55	18.24	19.36	25.00	16.47	1.28	6.67	11.11	35.90	13.30

Dari perhitungan model CAPM dengan menggunakan penduga kuadrat terkecil diperoleh penduga parameter CAPM beserta matriks varians kovariasnnya sebagai berikut :

Tabel 3. Nilai Parameter Model dan Matriks Kovarians

Perusahaan	Aphpa	Betha	Matriks Kovariance	
A	2.250	1.190	7.95	0.06
			0.06	0.1
B	2.450	1.660	10	0.07
			0.07	0.13
C	1.930	0.960	7.72	0.06
			0.06	0.1
D	1.970	1.000	10.85	0.08
			0.08	0.14
E	4.120	1.370	6.02	0.04
			0.04	0.08
F	2.230	0.010	1.05	0.01
			0.01	0.01
G	2.730	0.760	4.36	0.03
			0.03	0.06

Perusahaan	Aphpa	Betha	Matriks Kovariance	
H	6.330	0.220	12.17	0.09
			0.09	0.16
I	6.050	0.780	76.02	0.54
			0.54	0.97

Data di atas ini akan menjadi imput dalam analisis clustering parameter model regresi dengan tujuan menggerombolkan perusahaan-perusahaan berdasarkan risiko investasinya. Teknik penggerombolan yang digunakan adalah *kError* dengan menetapkan ada sebanyak dua gerombol yaitu perusahaan dengan risiko rendah dan perusahaan dengan risiko tinggi.

Proses perhitungan penggerombolan dengan k*Error* :

1. Menentukan keanggotaan gerombol pertama dengan teknik k-means untuk k=2
   Dari analisis gerombol dengan k-means diperoleh keanggotaan pengerombolan sebagai berikut :

Tabel 4. Penggerombolan Awal

Perusahaan	Gerombol
A	1
B	1
C	1
D	1
E	2
F	1
G	1
H	2
I	2

2. Iterasi Pertama :
   Menghitung centroid k*Error* dengan rumus : $\hat{\theta}_k = (\sum_{i \in S_k} \Sigma_i^{-1})^{-1} (\sum_{i \in S_k} \Sigma_i^{-1} x_i),\ k = 1,2$

   Diperoleh nilai :

$$\hat{\theta}_1 = \begin{bmatrix} 2.3257 \\ 0.3629 \end{bmatrix} dan\ \hat{\theta}_2 = \begin{bmatrix} 3.9036 \\ 0.7721 \end{bmatrix}$$

Menghitung jarak setiap objek ke centroid dengan rumus : $d_{ik} = (x_i - \hat{\theta}_k)^t \Sigma_i^{-1} (x_i - \hat{\theta}_k)$ diperoleh :

Tabel 6. Hasil Perhitungan Jarak Eucledean Iterasi 1

Objek	Jarak Euclidean		Gerombol 1	Gerombol 2	Jumlah Kuadrat (1)	Jumlah Kuadrat (2)
	Centroid 1	Centroid 2				
A	6.9	2.2	1	2	47.367	4.859
B	13.0	6.4	1	2	168.358	41.451
C	3.6	0.9	1	2	13.245	0.845
D	2.9	0.8	1	2	8.686	0.585
E	13.0	4.5	2	2	19.972	19.972
F	12.5	58.9	1	1	156.676	156.676
G	2.6	0.3	1	2	6.959	0.100
H	1.5	2.5	2	1	6.368	2.263
I	0.3	0.1	2	2	0.004	0.004
Total					427.63	226.75

Terlihat dari perhitungan jarak Euclidean untuk k*Error*, terjadi perubahan gerombol dimana dengan keanggotaan gerombol terlihat pada gerombol 2.

3. Iterasi Kedua :

   Karena terjadi perubahan keanggotaan penggerombolan maka dilakukan iterasi tahap dua dengan hasil sebagai berikut :

$$\hat{\theta}_1 = \begin{bmatrix} 2.5578 \\ 0.0236 \end{bmatrix} dan\ \hat{\theta}_2 = \begin{bmatrix} 2.7462 \\ 1.1076 \end{bmatrix}$$

   Jarak objek ke centroid :

Tabel 7. Hasil Perhitungan Jarak Eucledean Iterasi 2

Objek	Jarak Euclidean		Gerombol 3	Jumlah Kuadrat (3)
	Centroid 1	Centroid 2		
A	13.7	0.1	2	0.011
B	20.7	2.4	2	5.678
C	9.0	0.3	2	0.082
D	6.9	0.1	2	0.017

Objek	Jarak Euclidean		Gerombol 3	Jumlah Kuadrat (3)
	Centroid 1	Centroid 2		
E	22.8	1.1	2	1.250
F	0.1	120.8	1	0.013
G	9.0	2.0	2	4.077
H	1.3	6.3	1	1.816
I	0.7	0.3	2	0.074
Total				13.02

Perhatikan jarak Euclidean di atas terlihat tidak ada lagi keangotaan gerombol yang berpindah. Selain itu jumlah kuadar dari gerombol terakhir jauh lebih kecil dibandingkan dengan gerombol sebelumnya sehingga gerombol pada iterasi terakhir dinyatakan tepat.

Sehingga dapat disimpulkan bahwa Peruhaan A, B, C, D, E, G, dan I masuk dalam gerombol 2 sedangkan F dan H masuk ke dalam gerombol 1.

Jika diperhatikan gerombol 1 adalah perusahaan dengan tingkat resiko return yang paling rendah dan untuk gerombol 2 adalah perusahaan-peruhaan dengan resiko return tinggi.

KESIMPULAN

1. Metode *Error based clustering* khususnya k*Error* merupakan suatu teknik pengerombolan non-hirarkikal yang baik digunakan untuk data dengan kekeliruan pengukuran. Kekeliruan pengukuran ini muncul karena adanya suatu peroses penyederhanaan data.
2. Salah satu aplikasi dari k*Error* adalah untuk penggerombolan model parameter regresi dalam pembentukan *Capital Asset Pricing Model* (CAPM) memberikan hasil yang sedikit berbeda dengan k-means. Hasil simulasi yang dilakukan oleh Kumar(2007) menunjukkan kesalahan klasifikasi untuk k*Error* dalam pemodelan parameter regresi adalah 0%.

DAFTAR PUSTAKA

Banfield J.D. and A.E. Raftery. Model-based gaussian and non-gaussian clustering. *Biometrics*, 49:803–821, 1993

Celeux G. and G. Govaert. Normal parsimonious clustering models.*Pattern Recognition*, 28:781–793, 1995.

Kumar, M., Patel, *N*.R., and Woo, J., 2007. Clustering Data With Measurement *Error*, Computational Statistics & Data Analysis, Volume 51, Issue 12,6081-6101: 2007

Magidson j. and J.K. Vermut. Latent class models for clustering: A comparison with k-means. *Canadian Journal of Marketing Research*, 20:37–44, 2002.

Rice J.A.. *Mathematical Statistics and Data Analysis*. Duxbury Press,2nd edition, 1986.

Scott A. J. and M. J. Symons. Clustering methods based on likelihood ratio criteria. *Biometrics*, 27:387–397, 1971.

Zhang *N*.L.. Hierarchical latent class models for cluster analysis. *AAAI-02*, pages 230–237, 2002.